International Journal for Housing Science and Its Applications
Publish August 4, 2025. Volume 46, Issue 3 Pages 1924-1936

https://doi.org/10.70517/ijhsa463152

A Synergistic Study of Big Data Processing Methods Based on Quantitative Computational Analysis and Cloud Computing in the Digital Transformation of Enterprises

Qing Zhou[1] and Min Fu[1,*]
[1] The Economic Management School, Hunan Applied Technology University, Changde, Hunan, 415000, China
Corresponding authors: (e-mail: 13875105974@163.com).

Abstract This paper takes the listed high-tech manufacturing enterprises in Shanghai and Shenzhen A-shares from 2019 to 2024 as the research object, extracts the relevant panel data, and conducts descriptive statistics and correlation analysis on them. The panel data are utilized to establish the effect model based on the panel data, and the direct effect test is conducted. Through polynomial regression and response surface analysis, the synergistic mechanism between related variables is discussed. The empirical study shows that the digital transformation of enterprises and the development level of big data processing methods and cloud computing significantly promote enterprise innovation and development. And all of them showed a significant ($P<0.001$) positive effect in terms of effect test. The effects of synergistic interaction terms on both innovation quantity and innovation quality were significantly ($P < 0.01$ or $P < 0.05$) positive. Therefore, this synergistic mechanism is fully considered to promote the development of digital transformation innovation in enterprises.

Index Terms correlation analysis, panel data, response surface analysis, synergistic interaction

I. Introduction

In the context of the new era in which the global economy is increasingly moving towards a digital economy, enterprises are facing unprecedented pressure for change and opportunities for innovation. The rapid development of information technology, cloud computing, Internet of Things, artificial intelligence, blockchain, big data, augmented reality, virtual reality and other emerging technologies are reconfiguring the production mode, management mode and service pattern of enterprises [1]-[3]. These modern information technology tools not only change the business model, but also promote the digital transformation of the whole society [4]. In this environment, whether an enterprise can successfully implement digital transformation is not only related to its ability to adapt to the rapidly changing market demand, but also the key to determine whether it can be invincible in the fierce market competition [5], [6].

The digital transformation of an enterprise refers to the process in which the enterprise takes information technology as the core, makes comprehensive changes and optimization to the business, production and management processes, and makes the enterprise's business processes reach automation, intelligence and refined management [7]. This process involves not only innovation at the technological level, but also profound changes in the enterprise at multiple levels such as strategy, organization and culture [8], [9]. Through technological innovation, process optimization, and business model reshaping, enterprise digital transformation helps enterprises achieve breakthroughs in multiple dimensions, such as improving operational efficiency, enhancing innovation capability, and broadening market share [10]-[12]. In this process, enterprises also seek to provide practical references on how to effectively respond to transformation challenges and formulate reasonable strategies for their business development through detailed quantitative calculation and analysis [13], [14].

This paper collects panel data of listed high-tech manufacturing companies in Shanghai and Shenzhen A-shares from 2019-2024 through questionnaires. Descriptive statistics and correlation analysis were conducted on the panel data using State analysis software to gain insight into the correlation between each variable. Then the effect model based on the panel data was used to carry out direct effect and synergy effect tests to explore the big data processing methods and cloud computing, the impact of enterprise digital transformation on the number of enterprise innovation and its synergy mechanism. Based on the polynomial regression model, the response surface analysis of the synergistic relationship between the independent variables and the dependent variable is carried out, and the response surface diagram is drawn for visualization.

II. Polynomial regression and response surface analysis methods

Polynomial regression and response surface analysis can explore the complex relationship between two variables in more depth than the difference score method of objective measurement consistency research questions [15].

For b_0 (the intercept), b_n being the coefficients, whether the relationship between X_1, X_2 and Y (the relevant variables) is moderated by W, Edwards gives a specific test on his personal academic website:

$$\begin{aligned}Y &= b_0 + b_1X_1 + b_2X_2 + b_3X_1^2 + b_4X_1X_2 + b_5X_2^2 + b_6W + b_7WX_1 \\ &+ b_8WX_2 + b_9WX_1^2 + b_{10}WX_1X_2 + b_{11}WX_2^2 + \varepsilon_Y\end{aligned} \tag{1}$$

To further visualize the moderating effect, the response surface can be plotted. Equation (1) can be organized to become equation (2). If W is a continuous variable, then as suggested by Aiken and West, the mean of W plus or minus one standard deviation is substituted into equation (2) to obtain two equations, which can then be plotted on the response surface. If W is a dichotomous variable (e.g., $W = 0$ or $W = 1$), it is substituted into equation (2) to obtain two equations, respectively, which can then be plotted as response surface plots. Based on the Bootstrap results of the two equations and the response surface plots, the moderating effects in the three comparison scenarios can be determined. If we want to precisely compare whether the relationship between the independent variable and the outcome variable is significantly different under different conditions of high and low condition of the moderating variable, we can refer to Lee and Antonakis's suggestion of dividing the sample into two groups of high and low using the median of the moderating variable as a criterion and then comparing the coefficients of the high and low groups to see whether they are different or not:

$$\begin{aligned}Y &= (b_0 + b_6W) + (b_1 + b_7W)X_1 + (b_2 + b_8W)X_2 + (b_3 + b_9W)X_1^2 \\ &+ (b_4 + b_{10}W)X_1X_2 + (b_5 + b_{11}W)X_2^2 + \varepsilon_Y\end{aligned} \tag{2}$$

Some studies have analyzed the difference between the independent variable X_2 and the variable W as a moderating variable. The principle of testing the moderating effect is similar to that of equation (1), except that W is replaced by $(X_2 - W)$, which can be expressed in equation (3) through collation and simplification:

$$\begin{aligned}Y &= b_0 + b_1X_1 + b_2X_2 + b_3X_1^2 + b_4X_1X_2 + b_5X_2^2 + b_6W \\ &+ b_7WX_1 + b_8WX_2 + b_9X_1^2X_2 + b_{10}WX_1^2 + b_{11}X_1X_2^2 \\ &+ b_{12}WX_1X_2 + b_{13}X_2^3 + b_{14}WX_2^2 + \varepsilon_Y\end{aligned} \tag{3}$$

The steps to test the moderating effect of equation (3) are as follows: if there are control variables, first put the control variables into the equation to form model 1. then put the variables X_1, X_2, X_1^2, X_1X_2, and X_2^2 into the equation to form model 2 on the basis of model 1. and again put the variable W into the equation to form model 3 on the basis of model 2; Finally, on the basis of Model 3, the moderating terms WX_1, WX_2, $X_1^2X_2$, WX_1^2, $X_1X_2^2$, WX_1X_2, X_2^3, and WX_2^2 are put into the equation to form Model 4. If the DR^2 of Model 4 is significant, then it means that moderation of $(X_2 - W)$ exists. Theoretically, it is even possible to test whether there is moderation of the difference $(W_1 - W_2)$ between the two moderating variables W_1 and W_2, but in general, research does not involve such a complex test of the moderating effect.

In the study, it may also involve whether the relationship between X_1, X_2 and Y is affected by the mediating effect of M. There are two ways to test the mediating effect of M. The first way is to test it through two regression equations: firstly by regressing the mediating term M on the five independent variables X_1, X_2, X_1^2, X_1X_2 and X_2^2, and secondly by regressing the outcome variable Y on the five independent variables X_1, X_2, X_1^2, X_1X_2, X_2^2 and the mediating term M are regressed, see equation (4) and equation (5):

$$M = a_0 + a_1X_1 + a_2X_2 + a_3X_1^2 + a_4X_1X_2 + a_5X_2^2 + \varepsilon_M \tag{4}$$

$$Y = b_0 + b_1M + b_2X_1 + b_3X_2 + b_4X_1^2 + b_5X_1X_2 + b_6X_2^2 + \varepsilon_Y \tag{5}$$

Substituting equation (4) into equation (5) gives, after collation:

$$\begin{aligned} Y &= (b_0 + a_0 b_1) + (b_2 + a_1 b_1) X_1 + (b_3 + a_2 b_1) X_2 + (b_4 + a_3 b_1) X_1^2 \\ &+ (b_5 + a_4 b_1) X_1 X_2 + (b_6 + a_5 b_1) X_2^2 + (\varepsilon_Y + b_1 \varepsilon_M) \end{aligned} \tag{6}$$

In equation (6), the coefficient expressions for X_1, X_2, X_1^2, X_1X_2, and X_2^2 reflect the direct, mediated, and total effects of polynomial regression. The direct, mediated, and total effects can then be visualized by plotting the response surface plots for the three comparison scenarios.

Another way to test the mediation effect of polynomial regression is to multiply the values of X_1, X_2, X_1^2, X_1X_2, and X_2^2 by their coefficients in polynomial regression to form a block variable, and then use the block variable as the independent variable to test the mediation effect. This approach allows for the analysis of multiple mediation effects and the ability to compare the magnitude of the mediation effects, but does not explain the mediation effects in each of the three comparison scenarios.

III. Basic panel data modeling

The regression model has K number of explanatory variables, N number of cross-sections, and T time series span. Then the data vector for the i th cross-section is:

$$y_i = \begin{pmatrix} y_{i1} \\ y_{i2} \\ \vdots \\ y_{iT} \end{pmatrix}, X_i = \begin{pmatrix} x_{i1}^1 & x_{i1}^2 & \cdots & x_{i1}^K \\ x_{i2}^1 & x_{i2}^2 & \cdots & x_{i2}^K \\ \cdots & \cdots & \cdots & \cdots \\ x_{iT}^1 & x_{iT}^2 & \cdots & x_{iT}^K \end{pmatrix}, \alpha_i = \begin{pmatrix} \alpha_{i1} \\ \alpha_{i2} \\ \vdots \\ \alpha_{iT} \end{pmatrix}, \mu_i = \begin{pmatrix} \mu_{i1} \\ \mu_{i2} \\ \vdots \\ \mu_{iT} \end{pmatrix} \tag{7}$$

where: y_{it} - the value of the explanatory variable at cross section i and time t.

x_{it}^j - the value of the j th explanatory variable at cross section i and time t. $j = 1, 2, \cdots, K$. $i = 1, 2, \cdots, N$. $t = 1, 2, \cdots, T$.

It can be seen that the vector y is a matrix of $N \cdot T \times 1$; the vector X is a matrix of $N \cdot T \times K$. The vector α is a matrix of $N \cdot T \times 1$; the vector μ is a matrix of $N \cdot T \times 1$, and the vector β is a matrix of $N \cdot T \times K$. Of matrices. For such data, the panel data model can be expressed in the following matrix form:

$$y = \alpha + X\beta + \mu \tag{8}$$

A general linear panel data model can be expressed as:

$$y_{it} = \alpha_{it} + \beta_{it} x_{it} + u_{it} \text{ Which } i = 1, 2, \cdots, N; t = 1, 2, \cdots T \tag{9}$$

Assuming time-series parameter chirality, that is, the parameters do not vary with time, the above model can be expressed as:

$$y_{it} = \alpha_i + \beta_i x_{it} + u_{it} \; Which \; i = 1, 2, \cdots, N; t = 1, 2, \cdots T \tag{10}$$

At this point, the parameters α_i, β_i are individual period constants whose values are only affected by the cross-section cells. There can be two possibilities for the intercept and slope parameters, provided that the parameters do not vary over time.

One possibility is that the values of the parameters α_i, β_i, etc., change with different cross-sections, a situation that is modeled as a variable coefficient model for panel data.

Another possibility is that the regression slopes are the same but the intercepts are different, i.e., for different cross-sections β_i all take the same value, but the intercept α_i changes with different cross-sections. In this case, the model is a variable intercept model for panel data, which can be expressed as:

$$y_{it} = \alpha_i + \beta x_{it} + u_{it} \text{ Which } i = 1, 2, \cdots, N; t = 1, 2, \cdots T \tag{11}$$

Since the variable-intercept model is the most commonly used model for panel data analysis, the discussion on both the random effects and fixed effects models is based on the variable-intercept model:

$$y_{it} = \alpha_{it} + \beta x_{it} + u_{it} \text{ Which } i = 1, 2, \cdots, N; t = 1, 2, \cdots T \tag{12}$$

In the variable intercept model, α_{it} contains both time and cross-section effects, and the intercept term α_{it} in the panel data model can be further subdivided into aggregate effects, time effects, and cross-section effects, i.e:

$$\alpha_{it} = \alpha + \varepsilon_i + \lambda_t \tag{13}$$

where α represents the overall effect. ε_i represents the cross-section effect, which reflects the difference between different cross sections. The λ_t represents the period effect. For different α_{it} settings, it can be subdivided into three panel data models: mixed estimation model, fixed effects model and random effects model.

(1) Mixed estimation model

If there is no significant difference between individuals in terms of cross-section as well as time, then the panel data can be directly mixed together to estimate the parameters by ordinary least squares (OLS), which is a typical mixed estimation model [16]. The new model can be built as follows:

$$y_{it} = \alpha + \beta x_{it} + u_{it} \text{ Which } i = 1,2,\cdots,N; t = 1,2,\cdots T \tag{14}$$

where y_{it} is a vector of response variables, x_{it} is a vector of independent variables, α denotes the intercept term, and u_{it} is a vector of error terms. where α does not vary with i , t .

If the model is a mixed estimation model where the explanatory variables are uncorrelated with the error term, the mixed least squares estimates of the model parameters are consistent estimates for either $N \to \infty$ or $T \to \infty$.

(2) Fixed effects model

In panel data, if the intercept of the model is different for different cross-sections or different time series, the regression parameters can be estimated by adding dummy variables to the model, and such models are called fixed-effects models [17].

There are three types of fixed effects models, namely, individual fixed effects models, time fixed effects models and momentary individual fixed effects models. They are described below.

1) Individual fixed effects model

In panel data, if the intercept of the model is different for different cross-sections, such model is called individual fixed effect model, a strong assumption of individual fixed effect is $E(u_{it} / \alpha_i, X_{it}) = 0$, $i = 1,2,\cdots,N$ and its model can be expressed as:

$$y_{it} = \alpha_i + \beta x_{it} + u_{it} \text{ Which } i = 1,2,\cdots,N; t = 1,2,\cdots T \tag{15}$$

where y_{it} is the vector of response variables and x_{it} is the vector of independent variables. α_i is the random variable with i intercept terms for i individuals and u_{it} is the error term.

2) Time fixed effects model

In panel data, if for different time, the intercept of the model is different, call such model as time fixed effect model, its model can be expressed as:

$$y_{it} = \lambda_t + \beta x_{it} + u_{it} \text{ Which } i = 1,2,\cdots,N; t = 1,2,\cdots T \tag{16}$$

where y_{it} is the vector of response variables and x_{it} is the vector of independent variables. λ_t is the random variable with different intercept terms for different times, and u_{it} is the error term.

3) Individual time fixed effects model

In panel data, if for different time, not pass the individual, the intercept of the model is different, say such model is individual time fixed effect model, its model can be expressed as:

$$y_{it} = \alpha_i + \lambda_t + \beta x_{it} + u_{it} \text{ Which } i = 1,2,\cdots,N; t = 1,2,\cdots T \tag{17}$$

where y_{it} is a vector of response variables and x_{it} is a vector of independent variables. α_i is a random variable with different intercept terms for different cross-sections, and λ_t is a random variable also with different intercept terms for different times. u_{it} is the error term.

(3) Random effects model

For the panel data variable intercept model, if α_i is a random variable, its distribution is independent of X_{it} ; X_{it} is a matrix of $N \cdot T \times K$, and β is a matrix of $K \times 1$ with the same regression coefficient for different individuals. Y_{it} is the explanatory variable and u_{it} is the error term, and the model is called an individual random effects model when the following assumptions are met:

$$\alpha_i \sim iid(\alpha, \sigma_\alpha^2) \tag{18}$$

$$u_{it} \sim iid(0, \sigma_u^2) \tag{19}$$

Similarly time random effects models and individual time random effects models can be defined. The mixed least squares estimator of the random effects model is consistent but not valid.

IV. Theoretical analysis and research hypotheses

IV. A. Opportunities for Big Data and Cloud Computing in China

It is clear that information technology companies play a pivotal role in China's economic development. China's best IT companies, represented by Baidu, Alibaba and Tencent, have worked hard and made outstanding contributions to the rapid progress and market development of Big Data and Cloud Computing in China. Usually cloud services and applications are categorized in three different ways: architecture-as-a-service, platform-as-a-service, and software-as-a-service.

Cloud computing and cloud services have very significant advantages over traditional IT built environments, such as anytime, anywhere network connectivity, fast and elastic environment deployment, on-demand service model, and centralized and shared resource pools. Further by way of configuration, cloud computing and cloud services can be divided into private clouds (consisting of individual organization's private resources), public clouds (for the public masses), and hybrid clouds (combining the characteristics of the first two). The size of China's cloud computing and big data market is shown in Figure 1, where SaaS is the largest and the IaaS market is the fastest-growing of the cloud computing segments, with SaaS statistically more than doubling the combined size of the IaaS and PaaS markets. However, the market growth rate of IaaS and PaaS exceeds that of SaaS.

Among them, there is the highest growth rate of cloud services in Asia and especially in China. Similarly, sourced from research firm IDC, Big Data's global profits are forecasted to grow from $74 billion in 2016 to more than $3 trillion by 2024, with the Chinese region also being one of the fastest growing. The rapid development of Big Data and Cloud Computing locally in China provides valuable opportunities for Chinese companies to innovate in technology, create standards, and expand applications and services within the global market. This enables Chinese companies to compete and profitably compete with Western companies on a level playing field.

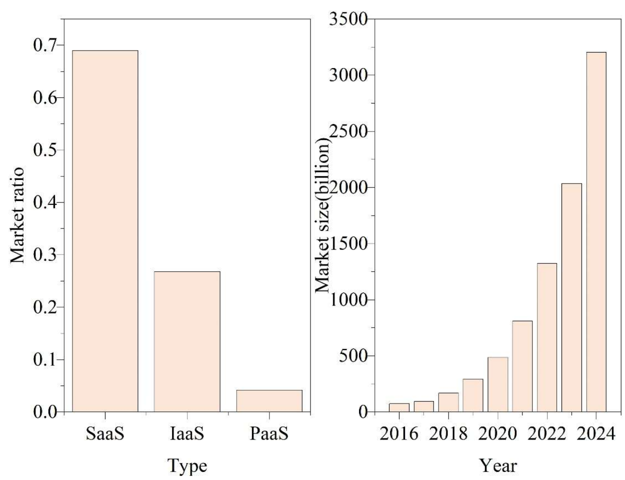

Figure 1: Large Numbers and cloud computing development

IV. B. Empirical analysis

IV. B. 1) Sample Selection and Data Sources

This study selects high-tech manufacturing enterprises listed in Shanghai and Shenzhen A-shares from 2019 to 2024 as the research object. Given that the “Made in China 2025” plan was first proposed in 2015, which aims to promote the digital transformation of the manufacturing industry, and considering that there is a certain lag effect in the implementation of the digital strategy, 2019 is chosen as the starting point of the study. Innovation quality (IQ, explanatory variable):The knowledge width of patents is adopted as the key indicator. Patent knowledge width reflects the complexity of the knowledge covered by the patent, which in turn reflects the quality of innovation from the dimensions of technology and value. Number of innovations (IN, explanatory variable): the total number of invention, utility model and appearance patent applications filed by enterprises in the same year is used as a proxy variable for the number of innovations, and the relevant data are obtained from the CNRDS database. Digital Transformation of Enterprises (DTT, Explanatory Variable): By mining and counting the keyword frequency in the annual reports of listed companies and standardizing them, it is finally used as a proxy variable for the digital transformation of enterprises, and the relevant data are obtained from the CNRDS database. Big data processing methods and the level of development of cloud computing (LDC, explanatory variables): cloud computing and big data are two sides of the same coin, cloud computing is the basis of big data, big data, providing storage space and access to big data channel of the growth of the driving force. Intermediary variables: regulatory transaction costs, information asymmetry. Control Variables: Return on Equity (ROE), Percentage of R&D Investment (PRD), Percentage of Net Cash Flow (Cashflow), Growth Rate of Operating Revenue (Growth), Enterprise Value (Tobin Q), Nature of Ownership (SOE), Number of Directors (Board), Percentage of Fixed Assets (Fixed), and Size of Firm (Size).

In order to guarantee the scientific validity and feasibility of the formal questionnaire before large-scale distribution, and to improve the confidence of sample data collection and analysis, this study conducted a small-scale pre-survey after the initial questionnaire design was completed. Offline, the middle and senior managers of the enterprises were invited to fill in the initial questionnaire mainly through visits, and online, the sample data were collected mainly by distributing a part of the questionnaire through the Creamdo online platform. A total of 116 questionnaires were collected, and 102 valid questionnaires were retained after review. First of all, the reliability analysis of the sample data of the pre-survey was carried out, mainly observing the value of Cronbach's alpha coefficient, which is closer to 1 the better, representing the more reliable data, the higher internal consistency, and more stability. All the scales of the questionnaire totaled 45 items, and the Cronbach's alpha coefficient value of the total scale was 0.811, indicating that the questionnaire scales used in this research have high internal consistency, as shown in Table 1.

Table 1: Pre-investigation reliability test results

Variable	Problem	Cronbach'sα
IQ	10	0.932
IN	11	0.844
DTT	7	0.892
LDC	7	0.853

Secondly, the validity test was conducted on the pre-study sample data, and the KMO value of the questionnaire scale was 0.731, which was greater than 0.7 and met the requirements, while the P value of Bartlett's spherical test was 0.000, which was less than 0.001, indicating that the scale was suitable for validation factor analysis. Then, the validation factor analysis was conducted on the pre-study data through Mplus software to further measure the discriminant validity of the scale and the overall fit of the model, and the results are shown in Table 2. Compared with the results of other models, the factor model fit was better, with a χ^2/df value of <3, indicating that the scale had good discriminant validity.

Table 2: Confirmatory factor analysis of pre-investigation

Factor combination	χ^2	df	χ^2/df	CFI	TLI	RMSEA
IQ	1378.526	940	1.56	0.786	0.772	0.072
IN	1607.694	943	1.79	0.652	0.635	0.075
DTT	1973.648	946	2.13	0.463	0.441	0.072
LDC	2247.573	951	2.41	0.314	0.289	0.078

The formal research lasts three months and is mainly collected through Creamdo online platform and Questionnaire Star applet. The questionnaire fillers need to be middle and senior managers of small and medium-sized manufacturing enterprises and be familiar with the overall situation of the enterprises' digital transformation strategy, industry-university-research cooperation, and governmental support, in order to ensure that high-quality sample data can be obtained. The questionnaires were distributed through three main channels: first, through the online platform through the limited conditions of the questionnaire distribution; second, through the alumni association, the group or personal social circle, in the form of QR code electronic questionnaires; and third, through the form of offline visits to the distribution of the questionnaires. The research time lasted three months, a total of 568 questionnaires were recovered, the sample data were reviewed and optimized, and the questionnaires whose quality did not meet the requirements were deleted, and finally 470 valid questionnaires were obtained, with an effective recovery rate of 82.75%. Descriptive statistical analysis of effective questionnaires, as shown in Table 3. Descriptive statistical analysis of effective questionnaires.

Table 3: Descriptive statistical analysis of samples

Variable	Classification	Frequency	Percentage
Position	Middle management	386	82.13%
	Senior management	84	17.87%
Enterprise age	Less than 3 years	80	17.02%
	3-6 years	115	24.47%
	7-9	174	37.02%
	Over 10	101	21.49%
Enterprise size	Under 100	98	20.85%
	100-499	87	18.51%
	500-1000	163	34.68%
	Over 1000	122	25.96%
Bachelor degree or above	Under 10%	70	14.89%
	11%-30%	242	51.49%
	31%-50%	79	16.81%
	Over 50%	79	16.81%
Type of enterprise ownership	State-owned enterprise	79	16.81%
	Private and private enterprises	242	51.49%
	Foreign enterprise	80	17.02%
	Other	69	14.68%
Industry	Food processing manufacturing	46	9.79%
	Manufacturing of electronic, computer and communication equipment	47	10.00%
	Medicine, biological products	43	9.15%
	Petroleum and chemical	36	7.66%
	General equipment manufacturing	40	8.51%
	Electrical machinery and equipment manufacturing	39	8.30%
	Transportation equipment manufacturing	27	5.74%
	Textile, clothing manufacturing	21	4.47%
	Tobacco manufacturing	31	6.60%
	Wood furniture	38	8.09%
	Metal and non-metallic products	30	6.38%
	Instrument manufacturing	45	9.57%
	Other	27	5.74%

IV. B. 2) Modeling

In order to test the impact of enterprise digital transformation and government digital governance on innovation "quality and quantity", this paper constructs a model to empirically test the impact of enterprise digital transformation and government digital governance on innovation "quality and quantity" with reference to existing studies:

$$\begin{aligned} IQ_{it} &= a_0 + a_1 DTT_{it} + a_m Control_{it} + \lambda_a + \lambda_h + \lambda_i + \varepsilon_{it} \\ IN_{it} &= a_0 + a_1 DTT_{it} + a_m Control_{it} + \lambda_a + \lambda_h + \lambda_i + \varepsilon_{it} \\ IQ_{it} &= a_0 + a_1 LDC_{it} + a_m Control_{it} + \lambda_a + \lambda_h + \lambda_i + \varepsilon_{it} \\ IN_{it} &= a_0 + a_1 LDC_{it} + a_m Control_{it} + \lambda_a + \lambda_h + \lambda_i + \varepsilon_{it} \end{aligned} \tag{20}$$

In this paper, the variables are centered on the basis of the introduction of the two and their interaction terms to explore the interaction effects, constructing a model as follows:

$$\begin{aligned} & IQ_{it} = a_0 + a_1 DTT_{it} + a_2 LDC_{it} + a_3 DTT_{ita} \times LDC_{it} + a_m Control_{it} \\ & + \lambda_a + \lambda_h + \lambda_i + \varepsilon_{it} \\ & IN_{it} = a_0 + a_1 DTT_{it} + a_2 LDC_{it} + a_3 DTT_{ita} \times LDC_{it} + a_m Control_{it} \\ & + \lambda_a + \lambda_h + \lambda_i + \varepsilon_{it} \end{aligned} \tag{21}$$

IV. B. 3) Descriptive statistics and correlation analysis

The descriptive statistics, as shown in Table 4, show that the minimum value of the number of innovations (IN) of enterprises is 1.125, and the maximum value is 6.58, indicating that there is a large variability in the number of innovations (IN) of enterprises. The mean value of enterprise innovation quality (IQ) is 0.482, and the standard deviation is 0.214, indicating that the overall level of innovation quality (IQ) is low, which is in line with the current basic situation that China's innovation quality (IQ) is generally not high. The mean value of enterprise digital transformation (DTT) degree is 2.078, and more than half of the enterprises in the sample have a DTT degree lower than the sample mean, indicating that a relatively high percentage of enterprises have a low DTT degree. This is similar to the conclusion that only 16% of Chinese enterprises with significant digital transformation (DTT) results were found in the 2021 China Enterprise Digital Transformation Index. From the analysis of the above results, the sample selection in this paper is representative. The mean value of the development level of big data processing methods and cloud computing (LDC) is 2.631, the minimum value is 1.396, and the maximum value is 3.335, indicating that there is a significant difference in this regard.

Table 4: Decriptive Statistics

	Observed number	Mean value	Standard deviation	Minimum value	Maximum value
IN	470	3.947	1.289	1.125	6.58
IQ	470	0.482	0.214	0	0.778
DTT	470	2.078	1.432	0	4.720
LDC	470	2.631	0.485	1.396	3.335
ASY	470	0.263	0.356	-0.298	1.469
COST	470	-0.0433	0.0224	-0.125	-0.0136
PRD	470	7.025	4.363	1.204	20.14
ROE	470	0.0768	0.0789	-0.198	0.221
Cashflow	470	0.0569	0.0514	-0.0416	0.178
Growth	470	0.154	0.220	-0.322	0.897
TobinQ	470	2.352	1.231	0.994	6.028
SOE	470	0.278	0.445	0	1
Board	470	2.097	0.189	1.611	2.458
Fixed	470	0.196	0.115	0.0345	0.479
Size	470	22.45	1.023	20.63	24.78

Before hypothesis testing, correlation analysis was performed on the variables and the results are shown in Figure 2, which shows the correlation coefficients between the variables, with a significant positive correlation between the digital transformation of the enterprise (DTT) and big data processing methods and the level of development of cloud computing (LDC) and the number of innovations (IN).

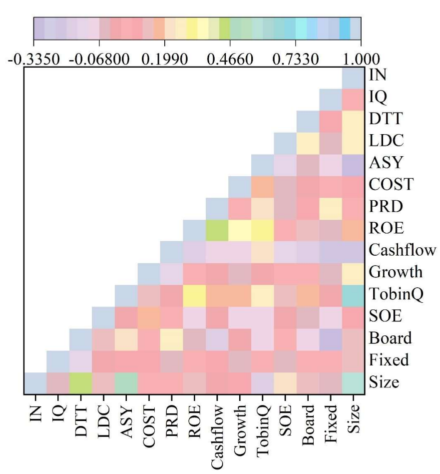

Figure 2: Correlation analysis

IV. B. 4) Direct effects

As shown in Table 5, this empirical result not only confirms the characteristic of "light on quality and heavy on quantity" in Chinese enterprise innovation, but also emphasizes the important role of big data and cloud computing in promoting the coordinated development of enterprise innovation quality (IQ) and quantity (IN). Therefore, in the process of promoting enterprise innovation and development in the future, more attention should be paid to the development level of big data processing methods and cloud computing (LDS) and enterprise digital transformation (DTT).

Table 5: Direct Effect Test

	(1)	(2)	(3)	(4)
	IQ	IN	IQ	IN
DTT	-0.008 (0.009)	0.054* (0.030)		
LDC			0.045*** (0.014)	0.170*** (0.042)
Fixed effect	Yes	Yes	Yes	Yes
N	470	470	470	470
R2	0.031	0.052	0.040	0.064
F	3.658	6.311	4.784	7.568

Table 6: Test of Interactive Synergy

	Interaction test		Synergy test	
	(1)	(2)	(3)	(4)
	IQ	IN	IQ	IN
DTT	-0.008 (0.009)	0.052* (0.032)	-0.006 (0.012)	0.062** (0.030)
LDC	0.052*** (0.014)	0.168*** (0.045)	0.056*** (0.016)	0.214*** (0.05)
DTT2			0.005 (0.006)	0.018 (0.019)
DTT×LDC	0.022* (0.009)	0.054* (0.035)	0.016* (0.009)	0.045* (0.038)
LDC2			0.028 (0.022)	0.174** (0.074)
Control variable	Control	Control	Control	Control
	Consistency line			
	X= y: slope		0.050*** (2.945)	0.284*** (5.027)
	X= y: curvature		0.053 (0.046)	0.223*** (3.089)
	Varying line			
	X= y: slope		-0.063*** (-3.268)	-0.154*** (0.009)
	X= y: curvature		0.016 (0.638)	0.138 (1.849)
N	470	470	470	470
R2	0.043	0.078	0.046	0.072
F	4.628	7.514	4.036	6.881

IV. B. 5) Synergy studies

(1) Interaction synergy effect analysis

The analysis of interaction synergistic effect is shown in Table 6, and the regression results of columns (1)~(2) respectively indicate that the interaction term of enterprise digital transformation and big data processing methods with cloud computing development level (DTT×LDC) has significantly positive coefficients on innovation quality (IQ) and quantity (IN), with estimated coefficients of 0.022 and 0.054 respectively, and with p-value less than 0.05, which indicates that there is a significant interaction effect between enterprise digital transformation (DTT) and there is a significant interaction between big data processing methods and cloud computing development level (LDC).

In order to better present the interaction effect, this paper further tests the synergy effect between government and enterprises. Column (3) is the synergistic effect on the innovation quality (IQ) of enterprises, and the regression results show that it indicates that when the level of enterprise digital transformation (DTT) and the level of big data processing methods and the development level of cloud computing (LDC) are consistent, the quality of innovation (IQ) will increase with the improvement of both of them. In the non-equilibrium state, the slope of the inconsistency line is significantly negative with an estimated coefficient of -0.063 and a p-value less than 0.01, and the curvature is not significant with an estimated coefficient of 0.016 and a p-value greater than 0.1. Column (4) is a synergistic effect of the number of firms' innovations (IN), and the regression results show that the number of firms' innovations (IN) increases with the level of the firms' digital transformation (DTT) and the level of the development of the big data processing methods and cloud computing (LDC). development level (LDC) shows a U-shaped development law, that is, the number of innovations (IN) decreases and then increases, and the slope of the inconsistency line is significantly negative in the unbalanced state, with an estimated coefficient of -0.154 and a p-value of less than 0.01, and the curvature is not significant with an estimated coefficient of 0.138 and a p-value of greater than 0. 1.

The response surface analysis of innovation quality is shown in Figure 3 The response surface is rotated clockwise along the inconsistency line, indicating that in the case of inconsistency, the higher the innovation quality (IQ) of an enterprise is when the level of development of big data processing methods and cloud computing (LDC) is higher than the level of development of digital transformation (DTT) of the enterprise, as compared to the case where the level of development of big data processing methods and cloud computing (LDC) is lower than the level of digital transformation (DTT) of the enterprise.

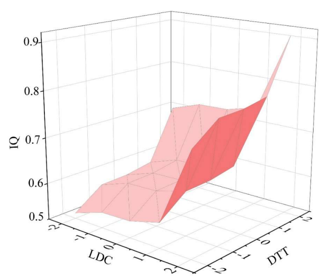

Figure 3: Response surface analysis of innovative quality

The response surface analysis of the number of innovations is shown in Figure 4, where the response surface is rotated clockwise along the inconsistency line, indicating that in the case of inconsistency, compared to the scenario where the level of development of big data processing methods and cloud computing (LDC) is lower than that of enterprise digital transformation (DTT), the number of innovations in the enterprise (IN) is higher when the level of development of big data processing methods and cloud computing (LDC) is higher than that of enterprise digital transformation (DTT) is higher.

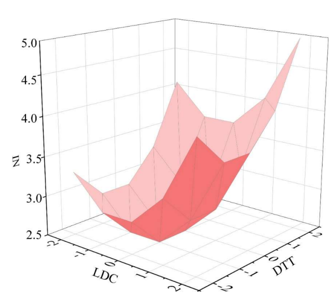

Figure 4: The response surface analysis of the innovation quantity

(2) Analysis of the influence mechanism of interaction synergy

In this paper, the block variable method is used to test the mediating effect of managerial transaction costs (COST) as shown in Table 7, and the estimated coefficients obtained from Column (1) are used as weights for linear weighted combinations, and the corresponding COST block variables are calculated for the test of the mediating effect of managerial transaction costs (COST). Enterprise digital transformation (DTT) and big data processing methods and the level of development of cloud computing (LDC) through the collaborative empowerment of organizations to manage transaction costs, to help enterprises to achieve the interconnection and interoperability of information, reduce the cost of management coordination of various departments of the enterprise, improve the efficiency of management decision-making, reduce the additional costs due to the solidification of policies, alleviate the enterprise financial constraints to improve the financial support of enterprise innovation activities, and ultimately promote Ultimately, it promotes the "quality and quantity" of enterprise innovation activities.

Table 7: The Impact Mechanism of Managerial Transaction Costs

	(1)	(2)	(3)	(4)	(5)	(6)
	COST	IQ	IN	COST	IQ	IN
COST block variable		1.263** (0.502)	6.714*** (1.638)	0.542*** (0.043)	0.659 (0.526)	5.278*** (1.785)
COST					1.152*** (0.335)	2.784*** (1.098)
DTT	0.006*** (0.002)					
LDC	0.013*** (0.002)					
DTT2	0.002*** (0.000)					
DTT×LDC	0.004*** (0.002)					
LDC2	-0.002 (0.003)					
Control variable	Control	Control	Control	Control	Control variable	Control
N	470	470	470	470	470	470
R2	0.563	0.026	0.029	0.242	0.036	0.033
F	5.978	2.645	13.487	6.052	2.638	13.544

Table 8: The Impact Mechanism of Information Asymmetry

	(1)	(2)	(3)	(4)	(5)	(6)
	ASY	IQ	IN	ASY	IQ	IN
ASY block variable		0.124*** (0.032)	0.522*** (0.093)	0.126*** (0.035)	0.112*** (0.030)	0.472*** (0.094)
ASY					0.063** (0.032)	0.332*** (0.093)
DTT	0.038*** (0.008)					
LDC	0.052*** (0.022)					
DTT2	0.012*** (0.006)					
DTT×LDC	0.009 (0.009)					
LDC2	0.074*** (0.025)					
Control variable	Control	Control	Control	Control	Control variable	Control
N	470	470	470	470	470	470
R2	0.312	0.036	0.040	0.305	0.042	0.051
F	10.314	2.784	15.463	10.847	2.638	11.472

In this paper, the mediating effect test of information asymmetry (ASY) using block variable approach is shown in Table 8. Enterprise digital transformation (DTT) and big data processing methods and cloud computing development level (LDC) through synergistic empowerment of enterprise information value, to help enterprises effectively disclose innovation information, coordination of innovation information sharing and confidentiality of the balance of the relationship between external stakeholders in a timely and accurate grasp of the value of enterprise innovation, the process and the risk of related information, and effectively alleviate the resource constraints of enterprise innovation, and ultimately promote the innovation activities of enterprises to "improve the quality and increase the quantity". It effectively alleviates the resource constraints of enterprise innovation, and ultimately promotes the "quality and quantity" of enterprise innovation activities.

V. Conclusion

This paper takes 396 listed companies in high-tech manufacturing industry as samples, and adopts quantitative analysis techniques such as correlation analysis, polynomial regression, response surface analysis and block variable analysis to analyze the interaction and synergistic mechanism of big data processing methods and cloud computing in the digital transformation of enterprises, and to guide and promote the digital transformation of enterprises. The research results show that:

(1) There is a significant positive correlation between enterprise digital transformation, big data processing methods and cloud computing development level, and the number of innovation-related variables. And big data and cloud computing significantly promote the coordinated development of enterprise innovation quality and quantity.

(2) The estimated coefficients of the interaction terms of enterprise digital transformation, big data processing methods and cloud computing development level on the quality and quantity of innovation are 0.022 and 0.054, respectively, with p-values less than 0.05, which indicates that there is a significant interaction effect between them.

Therefore, promoting the development level of big data and cloud computing technology enables the digital transformation of enterprises to be improved, which in turn promotes the overall development of their innovation quality and quantity.

Funding

A Key Teaching Reform Project of the Hunan Provincial Department of Education: Research on Approaches to Integrating the Revolutionary Spirit into Ideological Education within the Application-Oriented Undergraduate Statistics Curriculum Project ID (202401000193).

References

[1] You, B., & Xiao, X. (2023). Data encryption technology application in enterprise cost operation management based on cloud computing. Soft Computing, 1-13.

[2] Erasmus, J., Grefen, P., Vanderfeesten, I., & Traganos, K. (2018). Smart hybrid manufacturing control using cloud computing and the internet-of-things. Machines, 6(4), 62.

[3] Papadopoulos, T., Gunasekaran, A., Dubey, R., & Fosso Wamba, S. (2017). Big data and analytics in operations and supply chain management: managerial aspects and practical challenges. Production Planning & Control, 28(11-12), 873-876.

[4] Liu, P., & Wu, J. (2023). Can digital transformation enable the energy enterprises to achieve high-quality development?: An empirical analysis from China. Energy Reports, 10, 1182-1197.

[5] Liu, H., Wang, P., & Li, Z. (2021). Is there any difference in the impact of digital transformation on the quantity and efficiency of enterprise technological innovation? Taking China's agricultural listed companies as an example. Sustainability, 13(23), 12972.

[6] Cheng, L. (2022). Decision modeling and evaluation of enterprise digital transformation using data mining. Mobile information systems, 2022(1), 2380100.

[7] Chen, Q. A., Zhao, X., Zhang, X., Jiang, Z., & Wang, Y. (2024). Driving forces of digital transformation in Chinese enterprises based on machine learning. Scientific Reports, 14(1), 6177.

[8] Luo, X., & Chaiprasit, K. (2024). The Impact of Digital Transformation on Enterprise Performance. Rajapark Journal, 18(60), 64-83.

[9] Li, C., & Wang, Y. (2024). Digital transformation and enterprise resilience: Enabling or burdening?. Plos one, 19(7), e0305615.

[10] Chen, S., Moinzadeh, K., Song, J. S., & Zhong, Y. (2023). Cloud computing value chains: Research from the operations management perspective. Manufacturing & Service Operations Management, 25(4), 1338-1356.

[11] Bajdor, P. (2024). Evaluating Current and Future Impacts of Cloud Computing on Enterprise Operations: A Comparative Analysis. Procedia Computer Science, 246, 5185-5194.

[12] Asaul, A. A., Voynarenko, M., Yemchuk, L., & Dzhulii, L. (2020). New realities of the enterprise management system information support: Economic and mathematical models and cloud technologies. Journal of Information Technology Management, 12(3), 44-60.

[13] Peng, J., & Bao, L. (2023). Construction of enterprise business management analysis framework based on big data technology. Heliyon, 9(6).

[14] Zhang, W. J., Wang, J. W., & Lin, Y. (2019). Integrated design and operation management for enterprise systems. Enterprise Information Systems, 13(4), 424-429.

[15] Ningning Zhao,Diya Dou,Xiaohan Chen,Fumei Chen,Rui Luo,Xiaoqin Zhu & Guo Xing Xiang. (2025) .Exploring the congruence between perceived parent-teacher achievement goals and student academic outcomes: A study using polynomial regression with response surface analysis.. The British journal of educational psychology.

[16] Aibing Ji,Yu Cao,Jinjin Zhang & Qingqing Li. (2025) .SVR-based method for fixed effects interval-valued panel models. Communications in Statistics - Theory and Methods,54(8),2393-2411.

[17] Aubrey E. Parks,Anders K. Hansen & Brian W. Pogue. (2025) .Hybrid Monte Carlo model for efficient tissue Cherenkov emission estimation to assess changes from beam size, energy and incidence. Physica Medica,132,104956-104956.